ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

M.Sc. in Data Science

Course: Probability and Statistics for Data Analysis

Semester: Fall 2018

Instructor: Ioannis Vrontos (`vrontos@aueb.gr`)

Grader: Konstantinos Bourazas (`kbourazas@aueb.gr`)

Assignment 1

Deadline: 11 December 2018

1. Assume that A and B are events of the sample space S for which we know: $2P(A)-P(A')=3/5$, $P(B|A)=5/8$ and $P(A|B)=4/9$. Calculate the following probabilities:

(a) $P(A)$

(b) $P(A\cap B)$

(c) $P(B)$

(d) $P(A\cup B)$

(e) Are the events A and B independent?

2. Two players, A and B, alternatively and independently flip a coin and the first player to obtain a head wins. Assume player A flips first.

(a) If the coin is fair, what is the probability that player A wins?

(b) More generally assume that $P(head) = p$ (not necessarily $1/2$). What is the probability that player A wins?

(c) Show that $\forall p$ such that $0 < p < 1$, we have that $P(A\ wins) > 1/2$.

3. A telegraph signals "dot" and "dash" sent in the proportion $3 : 4$, where erratic transmission cause a dot to become dash with probability $1/4$ and a dash to become a dot with probability $1/3$.

(a) If a dash is received, what is the probability that a dash has been sent?

(b) Assuming independence between signals, if the message dot-dot was received, what is the probability distribution of the four possible messages that could have been sent?

4. Let X be a continuous random variable with pdf $f(x)$ and cdf $F(x)$. For a fixed number x_0 (such that $F(x_0) < 1$), define the function:

$$g(x) = \begin{cases} \dfrac{f(x)}{1 - F(x_0)} & \text{if } x \geq x_0 \\ 0 & \text{if } x < x_0 \end{cases}$$

Prove that $g(x)$ is a pdf (also known as hazard function).

5. Consider a telephone operator who, on the average, handles five calls every three minutes.

(a) What is the probability of no calls in the next minute?

(b) What is the probability of at least two calls in the next minute?

(c) What is the probability of at most two calls in the next five minutes?

6. Let $X_1, X_2, ..., X_n$ be a random sample form a $Gamma(\alpha, \beta)$ distribution. Find a two-dimensional sufficient statistic for (α, β).

7. One observation X is taken from a $N(0, \sigma^2)$ distribution.

(a) Find an unbiased estimate of σ^2.

(b) Find the maximum likelihood estimator (MLE) of σ^2.

8. Two random samples of size of $n = 10$ from a process producing bottles of water are gathered. The sample means are $\bar{x}_1 = 1000.42ml$ and $\bar{x}_2 = 999.58ml$ respectively. We assume that the data are normally distributed with $\sigma = 0.62$ (known).

(a) Provide a confidence interval for the mean of each subgroup in $\alpha = 0.05$ significance level.

(b) Test if the sample means of the subgroups are statistically equal in $\alpha = 0.05$ significance level.

(c) Test if $\bar{x}_1$ is statistically greater than $1Litre$ in $\alpha = 0.05$ significance level.